

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-155595

(43)Date of publication of application : 06.06.2000

(51)Int.Cl.

G10L 15/06  
 G10L 15/00  
 H04N 5/225  
 H04N 5/232  
 H04N 5/76

(21)Application number : 10-329386

(71)Applicant : CANON INC

(22)Date of filing : 19.11.1998

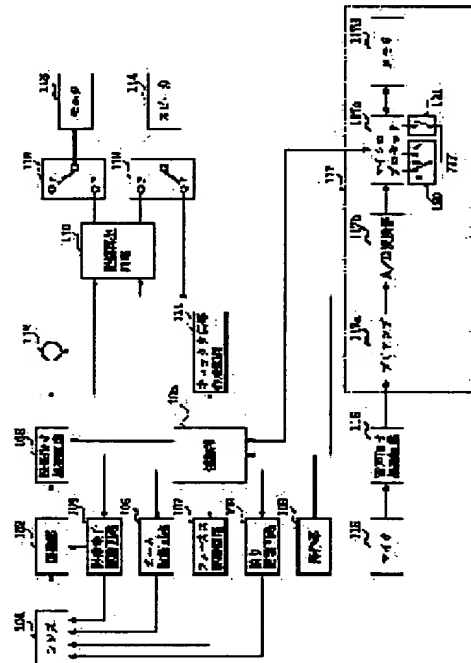
(72)Inventor : TAKEI HIROFUMI

## (54) IMAGE PICKUP DEVICE

## (57)Abstract:

**PROBLEM TO BE SOLVED:** To enhance reliability and response in speech recognition and to improve operability by providing a control means, a condition setting means for manually setting a recognizable condition for recognizing that an input speech signal is a prescribed speech signal in a speech recognition means and the like.

**SOLUTION:** A speech signal processing circuit 116 performs the well-known processing of a level adjustment, noise reduction, etc., to the speech signal from a microphone 15 to output it to a speech recognition circuit 17. The microprocessor 117c of the speech recognition circuit 117 applies the speech recognition processing to the speech signal from an A/D converter 117b to output the speech recognition result to a control part 105 through a data bus. Further, the microprocessor 117c operates according to a speech mode set by a switch 120. In such a manner, since this device is constituted so as to set a speech recognizable condition for a control function, the optimum recognizable condition is set, and the certainty and the responsiveness in the speech recognition are improved, that is, the operability is improved.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision  
of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office



## 【特許請求の範囲】

【請求項 1】 被写体を撮像すると共に、前記撮像された被写体光を電気信号に変換することにより得られた映像信号を処理する装置であって、音声信号を入力する入力手段と、前記入力された音声信号を認識する音声認識手段と、前記認識された音声信号に応じて前記装置の動作を制御する制御手段と、前記音声認識手段において、前記入力音声信号が所定の音声信号であると認識する為の認識条件をマニュアル操作により設定可能な条件設定手段とを有することを特徴とする撮像装置。

【請求項 2】 前記条件設定手段は、前記認識条件を対話的に入力することを特徴とする請求項 1 記載の撮像装置。

【請求項 3】 前記条件設定手段は、前記装置の複数種類の制御機能に対応して互いに異なる認識条件を設定可能であることを特徴とする請求項 1 記載の撮像装置。

【請求項 4】 前記音声認識手段は、参照音声信号と前記入力音声信号との相関の度合いを検出し、当該検出結果に基づいて前記入力音声信号を認識することを特徴とする請求項 1 記載の撮像装置。

【請求項 5】 前記音声認識手段は、前記検出された相関の度合いと所定の閾値とを比較し、当該比較結果に基づいて前記入力音声信号を認識することを特徴とする請求項 4 記載の撮像装置。

【請求項 6】 前記条件設定手段は、前記装置の複数の機能に対応づけて前記認識条件を設定可能であり、前記条件設定手段は前記各制御機能に対応づけて前記閾値を複数設定可能であることを特徴とする請求項 4 記載の撮像装置。

【請求項 7】 前記入力音声信号の認識結果に対応した所定の表示を行う表示手段を備えることを特徴とする請求項 1 記載の撮像装置。

【請求項 8】 前記装置は前記装置の複数種類の機能を制御可能であり、前記表示手段は、前記各制御機能を示すキャラクタを表示することを特徴とする請求項 7 記載の撮像装置。

【請求項 9】 前記装置は前記装置の複数種類の機能を制御可能であり、前記音声認識手段により前記入力音声信号が所定の音声信号であると認識された場合、前記表示手段は、前記各制御機能を示すキャラクタと前記音声認識手段により認識が成立した旨を示すキャラクタとを同一画面上に表示することを特徴とする請求項 7 記載の撮像装置。

【請求項 10】 前記音声認識手段により前記入力音声信号が所定の音声信号であると認識されたなかった場合、前記表示手段は、前記音声認識手段により認識が失敗した旨を表示すると共に、再度音声を入力する旨の勧告表示を行うことを特徴とする請求項 7 記載の撮像装置。

【請求項 11】 前記音声認識手段は、前記入力音声信

号を参照音声信号として登録する登録手段を備え、前記登録手段において前記入力音声信号が所定の音声レベルに達していない場合、前記表示手段は、前記入力音声信号の登録が失敗した旨を表示すると共に、再度音声を入力する旨の勧告表示を行うことを特徴とする請求項 7 記載の撮像装置。

【請求項 12】 前記制御手段は前記映像信号を得る撮像手段の撮像動作を制御し、前記条件設定手段は、撮像動作を制御するための音声信号の認識条件を設定することを特徴とする請求項 1 記載の撮像装置。

【請求項 13】 前記入力音声信号と前記映像信号を記録媒体に記録再生する記録再生手段を備え、前記制御手段は前記記録再生手段の動作を制御すると共に前記条件設定手段は、前記記録再生手段の記録再生機能を制御する為の音声信号の前記認識条件を設定することを特徴とする請求項 1 記載の撮像装置。

## 【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は撮像装置に関し、特に、音声で制御される装置に関するものである。

【0002】

【従来の技術】従来より、ビデオカメラで撮影したビデオ信号を処理する装置としてビデオカメラ一体型 VTR が知られている。そして、ビデオカメラ一体型 VTR では、小型なボディサイズに多くの機能を備えている。しかし、この種の装置においては、限られたスペースに多くの機能の操作部を設けている為、操作が煩雑になっている。

【0003】そこで、以上の様な問題点を解決する為、特開平 64-56428 号公報ではカメラの機能を制御する制御機構において、音声を入力する音声入力手段と、入力された音声信号を認識する音声認識手段と、認識結果に対応する制御内容に基づいてカメラの機能を制御する制御手段を有する音声入力カメラが提案されている。この発明では、音声によって、絞り、シャッター速度、動作モード等のカメラの機能を自由に設定することができる。

【0004】

【発明が解決しようとする課題】しかし、このような従来の装置では、記録、停止等の確実な音声認識が求められる動作である為に厳しい音声認識条件を必要とする機能においても、或いは、再生等の応答性のよさが求められる動作である為に寛容な音声認識条件を必要とする機能においても特に音声認識条件に差をつけていない。この為、音声認識条件が寛容すぎて確実な制御用音声の認識ができず、重要機能の誤動作が生じたり、音声認識条件が厳しすぎて動作の応答性が悪くなったりしていた。

【0005】本発明は、前述の如き問題を解決することを目的とする。

【0006】本発明の更に他の目的は、音声認識の確実

性と応答性を高め、操作性を向上させるところにある。

【0007】

【課題を解決するための手段】このような目的下において、本発明においては、被写体を撮像すると共に、前記撮像された被写体光を電気信号に変換することにより得られた映像信号を処理する装置であって、音声信号を入力する入力手段と、前記入力された音声信号を認識する音声認識手段と、前記認識された音声信号に応じて前記装置の動作を制御する制御手段と、前記音声認識手段において、前記入力音声信号が所定の音声信号であると認識する為の認識条件をマニュアル操作により設定可能な条件設定手段とを有する構成とした。

【0008】

【発明の実施の形態】以下、図面を用いて本発明の実施形態について詳細に説明する。

【0009】図1は本発明が適用されるカメラ一体型デジタルVTRの構成を第一の実施形態として示すブロック図である。

【0010】まず、通常の記録再生時の動作について説明する。図1において、レンズ101はレンズ、ズーム機構、アイリス等を含み、被写体光を撮像部102に導く。レンズ101はズーム駆動回路106より出力された制御信号に対応してズーム動作を行い、フォーカス駆動回路107より出力された制御信号に対応して焦点整合動作を行い、絞り駆動回路108より出力された制御信号に対応して絞り調整動作を行う。

【0011】撮像部102はCCDを使って被写体を撮像し、得られた被写体光を電気信号に変換し、映像信号処理回路103に出力する。また、撮像部102は、撮像素子駆動回路104により出力される制御信号に対応したシャッタースピードでCCDを駆動する。

【0012】映像信号処理回路103は撮像部102より出力された信号に対してクランプ、 $\gamma$ 補正、ホワイトバランス等、撮像された信号をテレビジョン信号に準拠した映像信号とするための周知の処理を施し、記録再生回路110及び加算器112に出力する。

【0013】ここで、加算器112は後述のキャラクタ信号作成回路111で作成されたキャラクタ信号を映像信号に加算し、スイッチ118に出力する。スイッチ118は映像信号記録時にr側端子に接続され、加算器112より出力された映像信号をモニタ113に出力する。

【0014】記録再生回路110は映像信号処理回路103より出力された映像信号に対してA/D変換、DCT、ハフマン符号化等周知の可変長符号化を用いて符号化することでその情報量を圧縮し、圧縮した映像信号に対して同期、ID等の付加、誤り訂正符号化等の記録する為に必要な処理を施し、記録に適した形態に変更する。更に、この様な処理を施した映像信号を磁気テープ上に

記録する。

【0015】この様に記録された映像信号は記録再生回路110により再生され、誤り訂正、可変長復号、逆DCT等の処理を施された後、スイッチ118に出力される。スイッチ118は映像信号再生時にp側端子に接続され、記録再生回路110からの再生信号をモニタ113に出力する。

【0016】一方、マイク115は音声を集音し、得られた音声は音声信号処理回路116に出力する。音声信号処理回路116はマイクからの音声信号に対してレベル調整、ノイズリダクション等の周知の処理を施し、記録再生回路110、スイッチ119及び音声認識回路117に出力する。記録再生回路110は音声信号処理回路116からの音声信号に対してA/D変換、シャプニング、誤り訂正等の周知の処理を施し、磁気テープ上に記録する。スイッチ119は音声信号記録時にr側端子に接続され、音声信号処理回路116より出力された音声信号をスピーカ114に出力する。音声認識回路117については後述する。

【0017】この様に記録された音声信号は記録再生回路110により再生され、誤り訂正、デシャプニング等の処理を施された後、スイッチ119に出力される。スイッチ119は音声信号再生時にp側端子に接続され、記録再生回路110からの再生信号をスピーカ114に出力する。

【0018】次に本形態における音声認識回路117の動作について説明する。一般的に音声認識装置には、話者を限定する特定話者用と話者を限定せず誰の音声でも認識する不特定話者用とがある。特定話者用は認識する語彙を予め発声させ、登録しておき、この登録したデータに基づいて認識を行うものである。この特定話者用の装置によれば、使用する特定の話者に認識系を設定できる為、システムの負荷が軽くなると共に高い認識率が期待でき、言語にも依存されにくい。

【0019】一方、不特定話者用は登録動作が不要なため、話者を選ばず、すぐに音声認識を動作させる簡便性はあるが、認識精度を上げる為には演算装置、メモリとも大規模なシステムが必要となる。本形態の如きカメラ一体型VTR場合、ユーザはほぼ一個人に限定されることがと小型かつ低コストを絶対条件とすること等から、特定話者であり、更に特定語彙を対象とする音声認識装置が適しているといえる。

【0020】この様な背景から本形態においては、特定話者に対応した音声認識システムを用いている。このため、認識する語彙をユーザに予め発声させ、その音声データを参照音声パターンとして登録する音声登録モードと、入力音声と参照音声パターンとのマッチングを行う音声認識モードとを備えている。

【0021】図1において音声モードの切換はスイッチ120で行い、スイッチ120は音声登録モードa及び

音声認識モードb、音声入力機能をOFFするポジションcの3ポジションを選択可能である。スイッチ121は登録、認識モード時にユーザが発声する音声を入力する際のトリガースイッチとなる音声入力スイッチである。

【0022】マイク115から音声信号処理回路116を介して音声認識回路117に入力された音声信号は、後述の如く音声認識回路117において音声認識処理を施され、音声認識回路117はその音声認識動作状況及び音声認識結果を制御部105に出力する。ここで音声認識回路117は、スイッチ120により音声登録モード若しくは音声認識モードに設定されている時のみ動作する。

【0023】ここで、制御部105はマイクロプロセッサを含み、音声認識回路117から出力された音声認識結果に応じて、撮像素子駆動回路104、ズーム駆動回路106、フォーカス駆動回路107、絞り駆動回路108、映像信号処理回路103、音声信号処理回路116を制御し、また、これら各回路の制御状態を示す情報をキャラクタ信号作成回路111に出力する。キャラクタ信号作成回路111は、制御部105より出力されるズーム値、絞り値等の各種情報及び操作メニューをキャラクタ表示する為のキャラクタ信号を作成し、加算器112に出力する。加算器112からの動作は前述の通りである。

【0024】また、制御部105は記録、再生、停止、モード設定等各種操作キーを有する操作部109の操作に応じて映像信号処理回路103、音声信号処理回路116、スイッチ118、スイッチ119等をも制御する。

【0025】次に、音声認識回路117の具体的な動作について説明する。音声信号処理回路116より入力された音声信号は、ブリアンプ117aにおいて所定ゲインで増幅され、A/D変換器117bに出力される。A/D変換器117bはブリアンプ117aより入力された音声信号をA/D変換し、マイクロプロセッサ117cに出力する。ここで、マイクロプロセッサ117cは音声認識に適したレベルとなるようにブリアンプ117aにフィードバック制御がかかるオートゲインコントロールを行っている。

【0026】マイクロプロセッサ117cはA/D変換器117bからの音声信号に対して後述の如く音声認識処理を行い、その音声認識動作状況と音声認識結果とをデータバスによって制御部105に出力する。また、マイクロプロセッサ117cはスイッチ120で設定される音声モードに応じて動作し、その動作については後述する。更に、マイクロプロセッサ117cは音声登録モード時に音声認識処理で得られた音声パターンと設定された閾値とをメモリ117dに出力する。

【0027】図2はマイクロプロセッサ117cが行う音声認識処理の流れを示すフローチャートである。こ

では、認識の対象となる標準パターンを作成し、入力音声と標準パターンの一致度、即ち両パターンの相関の度合いを判定することにより単語音声認識を行うパターンマッチング方式を用いている。

【0028】一致度が高い場合、即ち、標準パターンベクトルと入力音声パターンベクトルの距離が所定の閾値より小さい場合は、入力音声信号が音声登録モードで登録された音声信号と同一であると判断される。ここで、閾値は一致度の高低に対応しており、一致度の高低をユーザが設定することによって決まる。例えば、ユーザが一致度を高く設定すれば、認識条件が厳しくなり、正確に制御用音声を入力しなければ入力音声は認識されない。

【0029】図2において、音声認識モードが設定されているか否かを判別し(S201)、音声認識モードであった場合には音声認識モード処理を実行する(S202)。一方、S201で音声認識モードでなかった場合、音声登録モードが設定されているか否かを判別し(S203)、音声登録モードが設定されていた場合には音声登録モード処理を実行する(S204)。

【0030】次に各モードの具体的な動作について説明する。図3は図2の音声登録モード処理の動作を示すフローチャートである。スイッチ120がa端子側に設定されていると、音声登録モード処理をスタートし、図4(a)、図5(a)に示す様な機能表示画面を表示するべく制御部105に制御信号を出力する(S301)。そして、ユーザにより、表示された各機能の中から、一致度の設定及び対応する音声の登録を行うべき機能が選択されるのを待つ。図4(b)、図5(b)に示した如く、機能が選択されると(S302)、一致の度合いを選択するための表示を行うべく制御部105を制御し、ユーザにより一致度が選択されるのを待つ。そして、図4(c)、図5(c)に示した如く、選択した各機能に対して一致の度合いを設定する(S303)。この時、撮像部102による撮像動作は行われておらず、図4及び図5の背景は所定の映像となる。

【0031】図4(a)~(c)は一致度設定の表示例として、記録停止の場合を示すものである。操作部109のキー操作によりカーソルを移動させ、図示の如く、制御メニューより記録部制御を選択し、更に記録停止を選択し、一致度の高低を設定する。例えば、記録停止の様に重要な機能の場合は、誤動作防止の為に一致度を高く設定している。

【0032】図5(a)~(c)はAEの全自動モードの一致度設定の表示例である。図示の如く、各選択操作を対話的に行うことができ、選択された項目の確認操作を行うことができる。操作部109のキー操作によりカーソルの移動や入力決定動作を行い、一致度を設定する。AEの全自動モードは一般に多用されるので、一致度を低く設定している。

【0033】この様に機能及び一致度が決定すると、次に、選択されている機能に対応させる音声の入力及び登録を行う。S304において、スイッチ121がONされているかをマイクロプロセッサ117cが検知する。スイッチ121がONされていれば、制御部105から選択されている機能と、それに対応した一致度を読み込む。

【0034】S305においてユーザの入力音声の検出及び音声分析を行う。ここでユーザは選択された機能と入力音声とを対応させて登録すべく、機能名を発声する。例えば、図4(b)のように記録停止が選択されている場合、実際の使用状況を想定した発声を行う。即ち、登録する機能名として、選択された機能の語彙だけでなく、任意の語彙を登録時に発声することで独自の音声登録することができる。

【0035】発声した音声データはバンドパスフィルタ分析等の音声分析により分析パラメータベクトルの時系列に変換されると共に、音声の振幅パターン等から単語の開始点、終了点を決定し単語の切り出しを行う。

【0036】次にS306において特徴点抽出を行い、特徴点を特徴パラメータに変換する。ここでは得られたスペクトルのローカルピークを検出し、これらのみを2値化抽出する。得られた特徴パラメータに対して線形又は非線形の時間正規化処理を行い、音声パターンを生成する(S307)。

【0037】次にS308において音声パターンの信頼性判定が行われる。ここでは、生成された音声パターンが参照音声パターンとして登録するのに値するレベルに達しているかを判定する。信頼性が不十分であると判定されると、音声パターンは登録されず、S309において登録不可能であり、図6に示した様に、音声再入力を必要とする旨の勧告表示を行うべく制御部105に制御信号を出力する(S310)。制御部105はキャラクタ信号作成回路111を制御して、モニタ113にこの勧告表示を所定時間行い、S304において再度スイッチ121がONされるのを待つ。

【0038】一方、音声パターンの信頼性が十分であると判定されると、選択された一致度に対応した閾値を設定し、選択された機能、それに対応する音声パターン及び一致度に対応した閾値をそれぞれに対応させてメモリ117dに記憶させると共に登録完了の旨の表示を行うべく制御部105に制御信号を出力し、登録動作が完了する。ここで、一致度が高い場合は、入力音声パターンと参照音声パターンの距離が小さいということなので、閾値として大きい値を設定する。逆に、一致度が低い場合には、閾値として小さい値を設定する。

【0039】次に音声認識モード処理について説明する。図7は音声認識モード処理の動作を示すフローチャートである。スイッチ120がb端子側に設定されていると、音声認識モードに入る。S701においてスイッ

チ121がONになっていればS702においてユーザの入力音声の検出及び音声分析を行うと共に、他の操作キーの動作の受付を禁止する旨の信号を制御部105に出力する。特徴抽出S703、時間正規化S704については、前述の図3のS306、S307と同様の動作を行う。

【0040】次にS705において入力音声と参照音声パターンとのマッチング処理、即ち、時間正規化された入力音声パターンベクトルと参照音声パターンベクトルとの距離計算を行う。S705において計算された、入力音声パターンと各参照音声パターンとの距離の中で最小のものが認識された単語として判定される(S706)。

【0041】次にS707において、S706で判定された音声パターンに対応した一致度の閾値をメモリ117dより読みだし、S708において、この閾値を用いて入力音声パターンと参照音声パターンとの相関の度合い、即ち一致度の信頼性判定が行われる。ここでは、前述の如く計算された入力音声パターンと参照音声パターンとの距離をメモリ117dより読み出した閾値と比較する。もし、入力音声パターンと参照音声パターンとの距離が、閾値よりも大きければ認識信頼性がないと判定し、図8の様に、その旨の表示を行うべく制御部105に制御信号を出力すると共に、S701において再度スイッチ121がONされるのを待つ(S711)。また、距離が閾値よりも小さければ認識信頼性があると判定され、認識結果に対応してVTRを制御するとともに、モニタ113に認識結果に対応したモード表示をするべく制御部105に制御信号を出力する(S709)。図9～11は認識後のモード表示例を示したものである。モード表示を所定の時間行った後に、モード表示を中止し、通常の画面に戻す(S710)。

【0042】また、音声認識による機能制御を行いたくない場合は、スイッチ120を中止のポジションcにすれば、音声認識動作が中止され、機能制御も行われない。

【0043】また、本形態のVTRでは、予め各機能に対して異なる一致度を設定している。図12は音声認識される機能に対して予めVTRに設定されている一致度の例を示した表である。「記録停止」、「シャッター優先」等の重要な機能は、一致度が「高」に設定され、より確実に誤動作なく制御されるようになっている。「記録開始」、「全自動」等の使用頻度の高い機能は、一致度が「低」に設定され、応答性よく制御されるようになっている。ユーザは、これら各機能の中から所望の機能について前述の如く一致度を設定することができる。

【0044】このように、本形態においては、音声認識の一致度を機能別に設定することができる。従って、確実な制御用音声の認識が必要な機能には厳しい認識条件を、応答性のよさが必要である機能には寛容な認識条件

を設定可能である。そのため、各機能に対して最適な認識条件が設定され、操作性の向上をもたらすものである。

【0045】なお、本形態ではビデオカメラ一体型VTRに対して本発明を適用した場合について説明したが、これ以外にも、複数の機能を持つ記録再生装置に対して本発明を適用可能であり、同様の効果を有する。

【0046】

【発明の効果】以上説明したように、本発明によれば、制御機能に対して音声認識条件を設定可能な構成であるので、制御機能に最適な認識条件を設定することができ、音声認識の確実性と応答性の向上、即ち、操作性の向上をもたらすことができる。

【図面の簡単な説明】

【図1】本発明が適用されるビデオカメラ一体型VTRの構成を示す図である。

【図2】図1のマイクロプロセッサが行う音声認識処理のフローチャートである。

【図3】図2の音声登録モード処理の動作を示すフロー\*

\*チャートである。

【図4】記録停止に対する一致度設定のモニタ表示例である。

【図5】AEの全自動モードに対する一致度設定のモニタ表示例である。

【図6】登録すべき音声信頼性が信頼性不十分と判定された場合の音声再入力勧告表示の表示例である。

【図7】図2の音声認識モード処理の動作を示すフローチャートである。

【図8】一致度判定時に信頼性不十分と判定された場合の音声再入力勧告表示の表示例である。

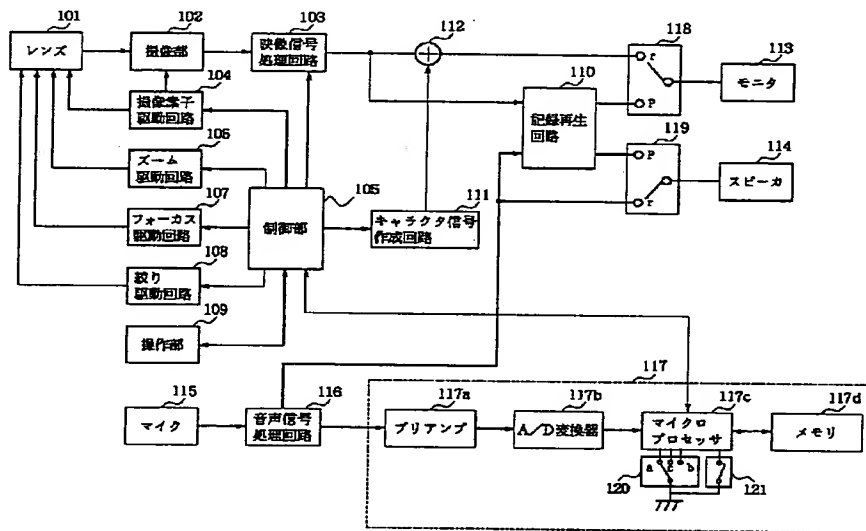
【図9】AEの全自動モードに対する一致度判定時に信頼性十分と判定された場合のモニタ表示例である。

【図10】記録開始に対する一致度判定時に信頼性十分と判定された場合のモニタ表示例である。

【図11】記録停止に対する一致度判定時に信頼性十分と判定された場合のモニタ表示例である。

【図12】音声認識される制御機能と各制御機能に対して設定された一致度の例である。

【図1】

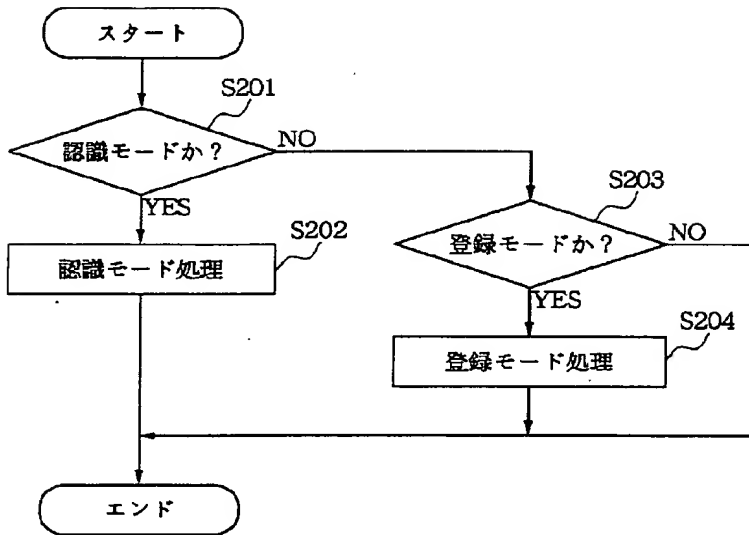


【図12】

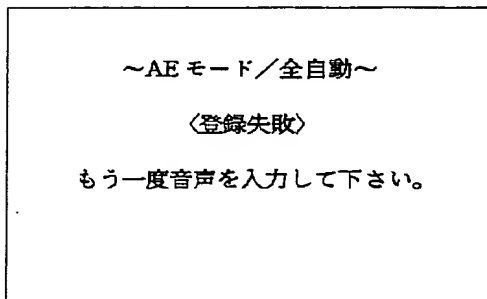
制御機能	一致度	AEモード	一致度	AFモード	一致度
記録開始	低	全自動	低	オート	低
記録停止	高	シャッター優先	高	マニュアル	高
画像消去	高	絞り優先	高		
早送り	中	スポーツ	中		
巻き戻し	中				
再生停止	低				



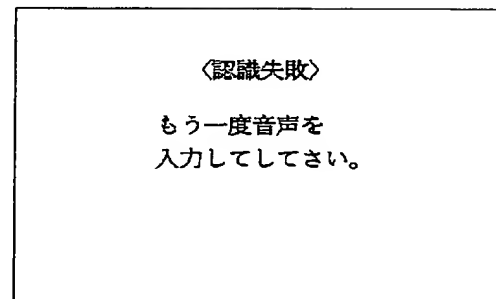
【図2】



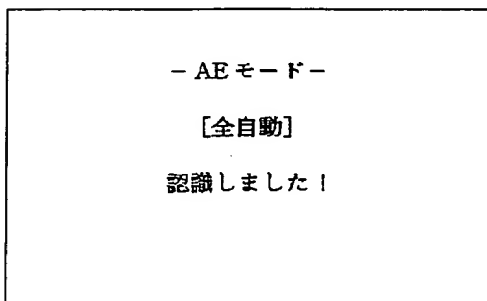
【図6】



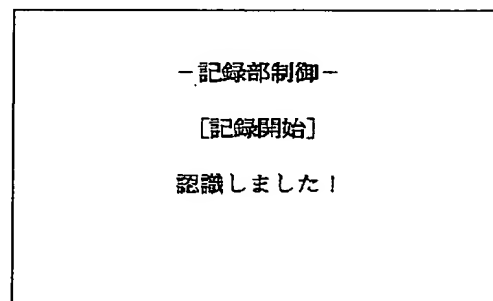
【図8】



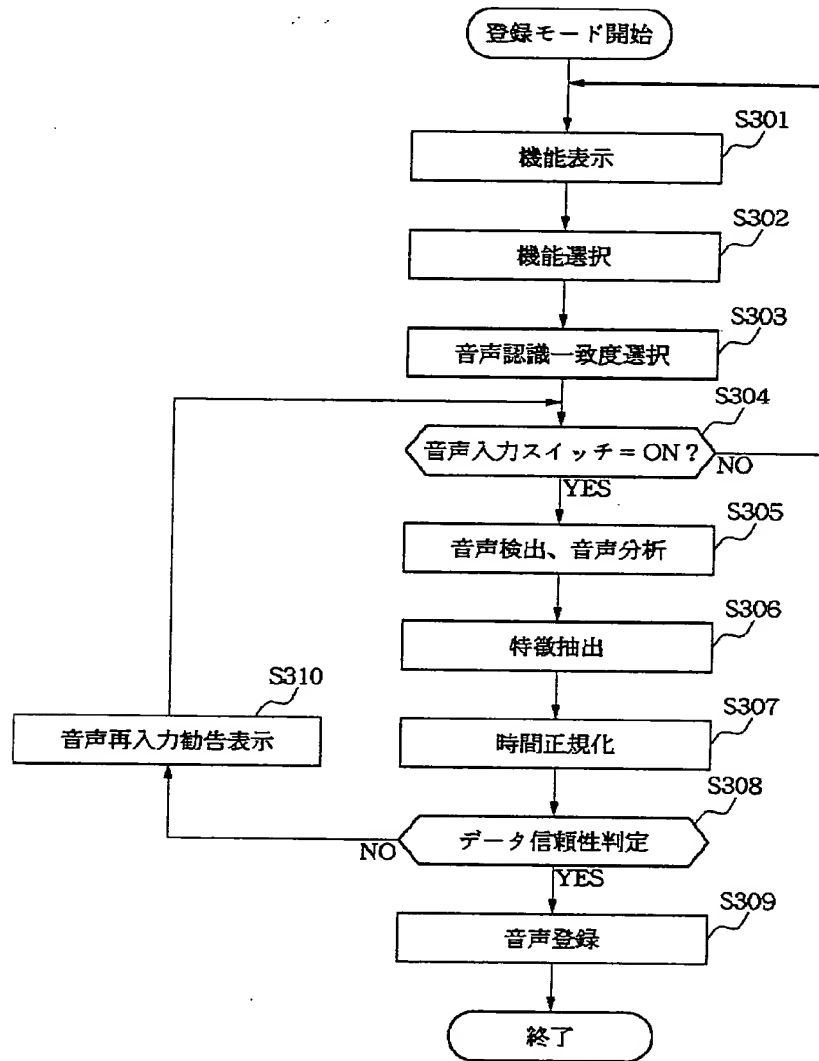
【図9】



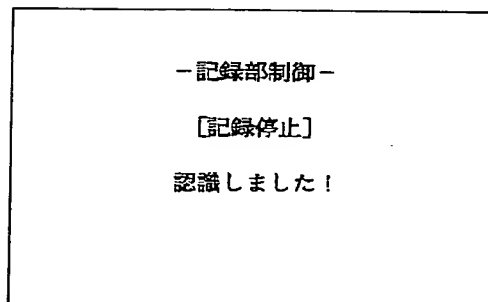
【図10】



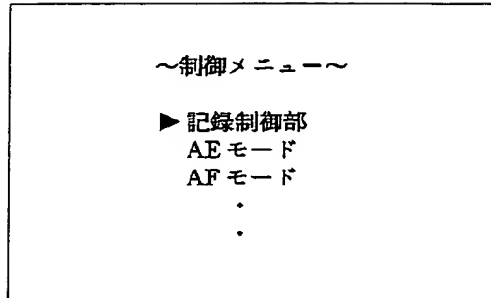
【図3】



【図11】

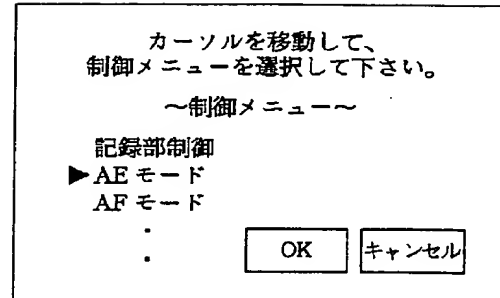


【図4】

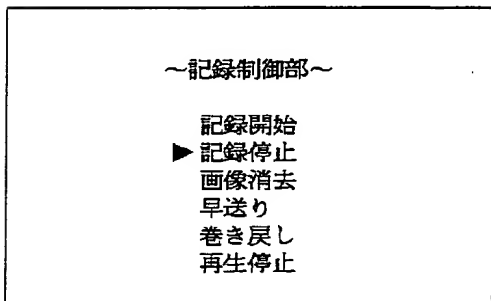


(a)

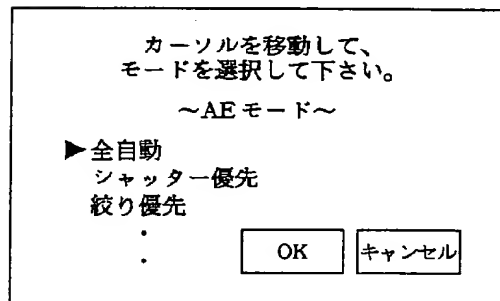
【図5】



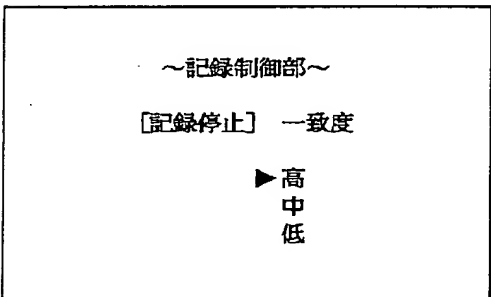
(a)



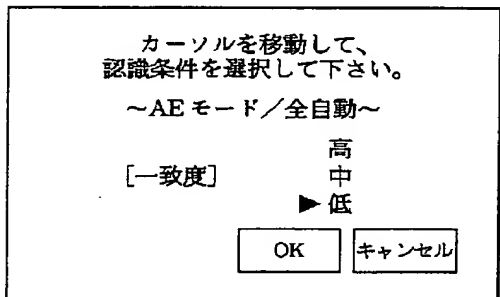
(b)



(b)

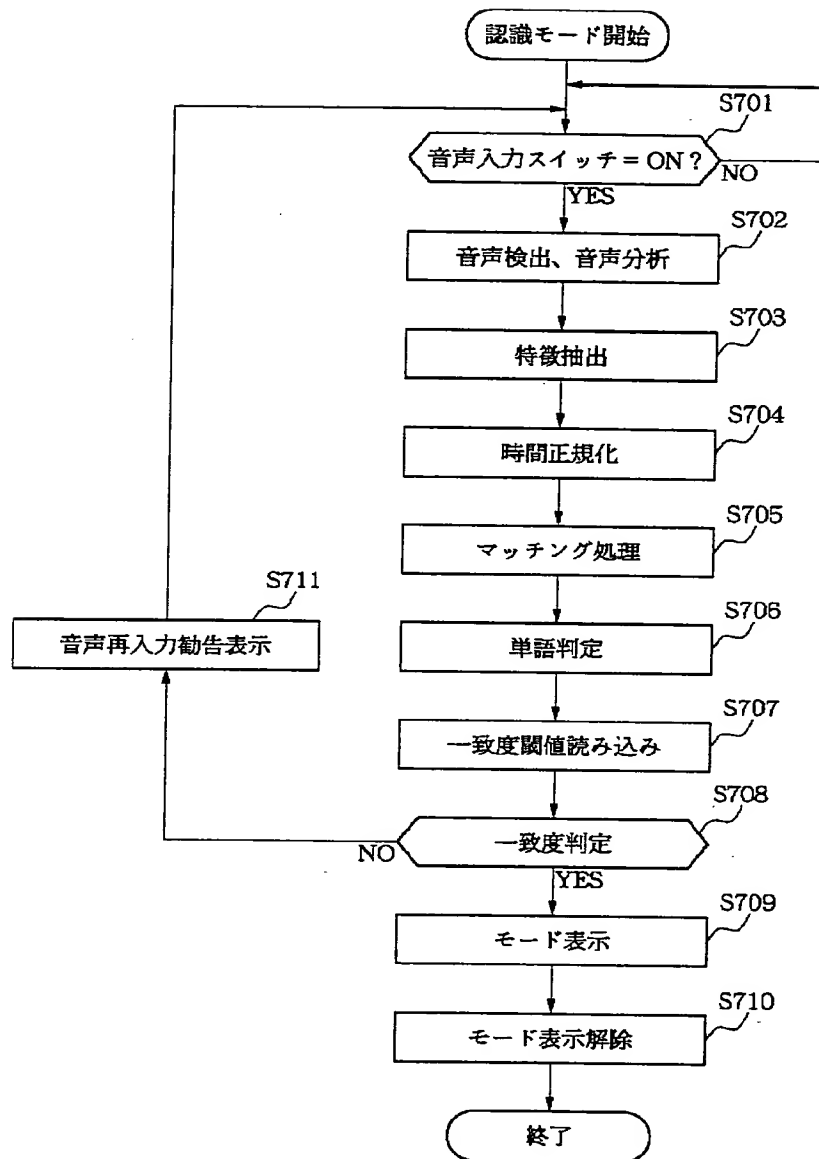


(c)



(c)

【図7】



フロントページの続き

(51)Int.Cl.<sup>7</sup>

識別記号

F I  
G 1 0 L 3/00

5 5 1 G

キーワード(参考)